

1/5/4 (Item 4 from file: 351)  
DIALOG(R) File 351:Derwent WPI  
(c) 2006 The Thomson Corp. All rts. reserv.

013901536 \*\*Image available\*\*  
WPI Acc No: 2001-385749/ 200141  
XRPX Acc No: N01-283332

**Speech translation system controls condition of robot, based on translated speech signal**

Patent Assignee: SONY CORP (SONY )  
Number of Countries: 001 Number of Patents: 001  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 2001117752	A	20010427	JP 99293876	A	19991015	200141 B

Priority Applications (No Type Date): JP 99293876 A 19991015

**Patent Details:**

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
JP 2001117752	A	14	G06F-003/16	

**Abstract (Basic): JP 2001117752 A**

NOVELTY - Speech recognition unit (1) recognizes user's voice which is input via input unit (3). The machine translation unit (2) translates recognized voice. The interface unit interfaces speech recognition unit and machine translation unit. Based on output of interface unit, a control unit (6) controls condition of robot.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

- (a) Information processing procedure;
- (b) Information processing program.

USE - Speech translation system provided with robot control function.

ADVANTAGE - Smooth communication is performed, as speaker's emotion is correctly communicated by controlling condition of robot based on translated output.

DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of speech translation system. (Drawing includes non-English language text).

Speech recognition unit (1)  
Machine translation unit (2)  
Input unit (3)  
Control unit (6)  
pp; 14 DwgNo 1/15

Title Terms: SPEECH; TRANSLATION; SYSTEM; CONTROL; CONDITION; ROBOT; BASED; TRANSLATION; SPEECH; SIGNAL

Derwent Class: T01

International Patent Class (Main): G06F-003/16

International Patent Class (Additional): G06F-003/00; G06F-017/28

File Segment: EPI

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-117752  
(43)Date of publication of application : 27.04.2001

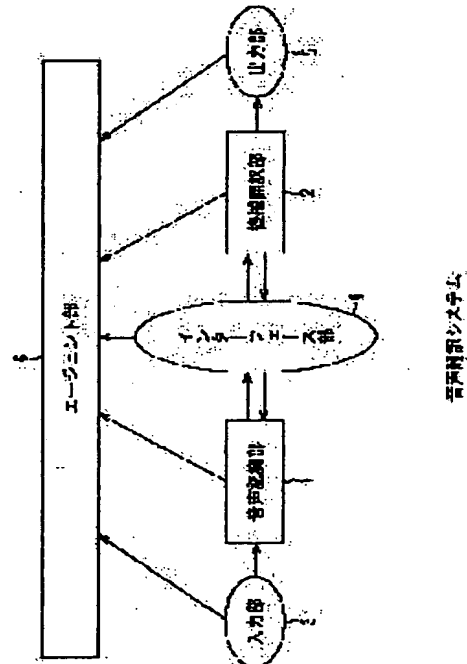
(51)Int.Cl. G06F 3/16  
G06F 3/00  
G06F 17/28

(21)Application number : 11-293876 (71)Applicant : SONY CORP  
(22)Date of filing : 15.10.1999 (72)Inventor : TSUTSUMI HIRONAGA

**(54) INFORMATION PROCESSOR, INFORMATION PROCESSING METHOD AND RECORDING MEDIUM**

**(57)Abstract:**

**PROBLEM TO BE SOLVED:** To attain smooth communication.  
**SOLUTION:** User's voice is supplied to a voice recognition part 1 through an input part 3, the voice is recognized and the recognized voice is supplied to a machine translation(MT) part 2 through an interface part 4. The MT part 2 applies machine translation to the voice recognition result of the user's voice and supplies the translated result to an output part 5. The output part 5 generates a synthetic sound corresponding to the translation result of the MT part 2. An agent part 6 controls characters displayed as images or the state of a robot to be a practical object on the basis of information expressing user's feeling and outputted from the voice recognition part 1, the MT part 2, the input part 3, the interface part 4 and the output part 5.



## LEGAL STATUS

[Date of request for examination]	01.03.2006
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	
[Date of registration]	
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	
[Date of extinction of right]	

52 512

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2001-117752  
(P2001-117752A)

(43) 公開日 平成13年4月27日 (2001.4.27)

(51) Int.Cl.	識別記号	F I	キーワード(参考)
G 0 6 F 3/16	3 2 0	G 0 6 F 3/16	3 2 0 H 5 B 0 9 1
3/00	6 0 1	3/00	6 0 1 5 E 5 0 1
17/28		15/38	V

審査請求 未請求 請求項の数14 O L (全 14 頁)

(21) 出願番号 特願平11-293876

(22) 出願日 平成11年10月15日 (1999. 10. 15)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(72) 発明者 包 洪長

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(74) 代理人 100082131

弁理士 稲本 義雄

Fターム(参考) 5B091 AA04 BA03 CB12 CB24 CB32

CC01

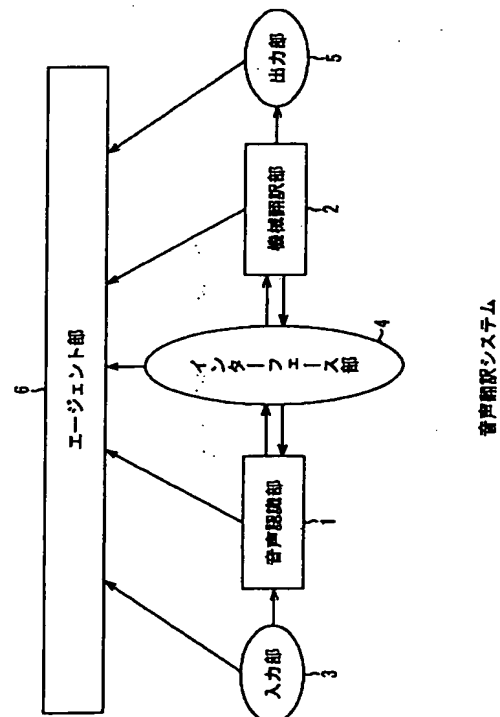
5E501 AC37 BA15 CB15 FA14 FA32

(54) 【発明の名称】 情報処理装置および情報処理方法、並びに記録媒体

(57) 【要約】

【課題】 円滑なコミュニケーションを図ることができるようにする。

【解決手段】 ユーザの音声は、入力部3を介して音声認識部1に供給され、音声認識され、インターフェース部4を介して、機械翻訳部2に供給される。機械翻訳部2では、ユーザの音声の音声認識結果が機械翻訳され、出力部5に供給される。出力部5では、機械翻訳部2による翻訳結果に対応する合成音が生産される。一方、エージェント部6は、音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5からのユーザの感情を表す情報に基づいて、画像として表示されるキャラクターや、実際の物体としてのロボットの状態を制御する。



## 【特許請求の範囲】

【請求項1】 音声を認識する音声認識手段と、前記音声認識手段による音声認識結果を翻訳する翻訳手段と、

前記音声認識手段と翻訳手段との間のインターフェースとして機能するインターフェース手段とを有する音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態を制御する制御手段を備えることを特徴とする情報処理装置。

【請求項2】 前記制御手段は、前記インターフェース手段から得られる情報に基づいて、前記オブジェクトの状態を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項3】 前記制御手段は、前記音声認識手段または翻訳手段から得られる情報に基づいて、前記オブジェクトの状態を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項4】 前記制御手段は、前記音声認識手段が認識する音声の言語、または前記翻訳手段による翻訳結果の言語の言語的な特徴に基づいて、前記オブジェクトの状態を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項5】 前記オブジェクトは、画像として表示されるキャラクタ、音声、またはロボットであることを特徴とする請求項1に記載の情報処理装置。

【請求項6】 前記制御手段は、画像の表示、または音声の出力を行うことが可能となっていることを特徴とする請求項1に記載の情報処理装置。

【請求項7】 前記制御手段は、前記音声認識手段による音声認識結果、または前記翻訳手段による翻訳結果の補正を行うことが可能となっていることを特徴とする請求項1に記載の情報処理装置。

【請求項8】 前記制御手段は、前記音声認識手段が音声認識を行うための知識、または前記翻訳手段が翻訳を行うための知識に基づいて、前記オブジェクトの状態を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項9】 前記制御手段は、前記音声認識手段が音声認識を行うための知識、または前記翻訳手段が翻訳を行うための知識に基づいて、前記オブジェクトを成長させることを特徴とする請求項1に記載の情報処理装置。

【請求項10】 前記制御手段は、前記音声翻訳装置から得られる情報に基づいて、前記オブジェクトの感情表現を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項11】 前記制御手段は、前記音声翻訳装置から得られる、前記音声の発話者の感情に基づいて、前記オブジェクトの感情表現を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項12】 前記制御手段は、前記音声認識手段に

よる音声認識結果の内容的な特徴、または前記翻訳手段による翻訳結果の内容的な特徴に基づいて、前記オブジェクトの感情表現を制御することを特徴とする請求項1に記載の情報処理装置。

【請求項13】 音声を認識する音声認識手段と、前記音声認識手段による音声認識結果を翻訳する翻訳手段と、

前記音声認識手段と翻訳手段との間のインターフェースとして機能するインターフェース手段と

10 有する音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態を制御する制御ステップを備えることを特徴とする情報処理方法。

【請求項14】 コンピュータが実行するプログラムが記録されている記録媒体であって、

音声を認識する音声認識手段と、

前記音声認識手段による音声認識結果を翻訳する翻訳手段と、

20 前記音声認識手段と翻訳手段との間のインターフェースとして機能するインターフェース手段とを有する音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態を制御する制御ステップを備えるプログラムが記録されていることを特徴とする記録媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、情報処理装置および情報処理方法、並びに記録媒体に関し、特に、例えば、入力された音声を認識し、その認識結果を翻訳する音声翻訳システムを用いて、円滑なコミュニケーションを図ることができるようにする情報処理装置および情報処理方法、並びに記録媒体に関する。

## 【0002】

【従来の技術】例えば、日本語と英語等の、異なる言語による発話を行うユーザどうしがコミュニケーションを図るためのツールとして、音声認識装置、翻訳装置、および音声合成装置から構成される音声翻訳システムがある。音声翻訳システムでは、音声認識装置において、例えば、日本語の発話が音声認識され、翻訳装置において、その音声認識結果が英語に翻訳される。そして、音声合成装置において、その翻訳結果が、合成音で出力される。また、音声認識装置において、英語の発話が音声認識され、翻訳装置において、その音声認識結果が日本語に翻訳される。そして、音声合成装置において、その翻訳結果が、合成音で出力される。従って、英語の話者（ユーザ）は、日本語の話者の発話を、英語で聞くことができ、また、日本語の話者は、英語の話者の発話を、日本語で聞くことができ、相互に、相手の発話を理解して対話を行うことができる。

## 【0003】

【発明が解決しようとする課題】ところで、上述のような音声翻訳システムでは、例えば、ユーザの発話の翻訳

結果に対応する合成音には、そのユーザの感情等が反映されない。このため、ユーザの意図が、相手のユーザに正確に伝わらないことがあり、円滑なコミュニケーションを図ることが困難な場合がある。

【0004】本発明は、このような状況に鑑みてなされたものであり、ユーザどうしが、円滑なコミュニケーションを図ることができるようにするものである。

【0005】

【課題を解決するための手段】本発明の情報処理装置は、音声認識する音声認識手段と、音声認識手段による音声認識結果を翻訳する翻訳手段と、音声認識手段と翻訳手段との間のインターフェースとして機能するインターフェース手段とを有する音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態を制御する制御手段を備えることを特徴とする。

【0006】制御手段には、インターフェース手段から得られる情報に基づいて、オブジェクトの状態を制御させることができる。

【0007】また、制御手段には、音声認識手段または翻訳手段から得られる情報に基づいて、オブジェクトの状態を制御させることができる。

【0008】さらに、制御手段には、音声認識手段が認識する音声の言語、または翻訳手段による翻訳結果の言語の言語的な特徴に基づいて、オブジェクトの状態を制御させることができる。

【0009】また、オブジェクトは、画像として表示されるキャラクタ、音声、またはロボットとすることができる。

【0010】制御手段には、画像の表示、または音声の出力を行わせることができる。

【0011】さらに、制御手段には、音声認識手段による音声認識結果、または翻訳手段による翻訳結果の補正を行わせることができる。

【0012】また、制御手段には、音声認識手段が音声認識を行うための知識、または翻訳手段が翻訳を行うための知識に基づいて、オブジェクトの状態を制御させることができる。

【0013】さらに、制御手段には、音声認識手段が音声認識を行うための知識、または翻訳手段が翻訳を行うための知識に基づいて、オブジェクトを成長させることができる。

【0014】また、制御手段には、音声翻訳装置から得られる情報に基づいて、オブジェクトの感情表現を制御させることができる。

【0015】また、制御手段には、音声翻訳装置から得られる、音声の発話者の感情に基づいて、オブジェクトの感情表現を制御させることができる。

【0016】さらに、制御手段には、音声認識手段による音声認識結果の内容的な特徴、または翻訳手段による翻訳結果の内容的な特徴に基づいて、オブジェクトの感

情表現を制御させることができる。

【0017】本発明の情報処理方法は、音声認識する音声認識手段と、音声認識手段による音声認識結果を翻訳する翻訳手段と、音声認識手段と翻訳手段との間のインターフェースとして機能するインターフェース手段とを有する音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態を制御する制御ステップを備えることを特徴とする。

【0018】本発明の記録媒体は、音声認識する音声認識手段と、音声認識手段による音声認識結果を翻訳する翻訳手段と、音声認識手段と翻訳手段との間のインターフェースとして機能するインターフェース手段とを有する音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態を制御する制御ステップを備えるプログラムが記録されていることを特徴とする。

【0019】本発明の情報処理装置および情報処理方法、並びに記録媒体においては、音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態が制御される。

【0020】

【発明の実施の形態】図1は、本発明を適用した音声翻訳システム（システムとは、複数の装置が論理的に集合した物をいい、各構成の装置が同一筐体中にあるか否かは問わない）の一実施の形態の構成例を示している。

【0021】この音声翻訳システムでは、例えば、日本語による音声が入力されると、その音声は英語に翻訳されて出力され、また、英語による音声が入力されると、その音声は日本語に翻訳されて出力されるようになっており、これにより、日本語のユーザ（話者）と、英語のユーザとが、対話を行うことができるようになってい

る。

【0022】即ち、音声認識部1は、入力部3からの音声データを受信し、その音声データを用いて、ユーザが発話した音声の音声認識する。そして、その音声認識結果としてのテキスト、その他付随する情報を、インターフェース部4やエージェント部6に出力する。

【0023】機械翻訳部2は、音声認識部1から、インターフェース部4を介して供給される音声認識結果を解析し、入力された音声を、その音声の言語以外の言語に機械翻訳し（本実施の形態では、日本語は英語に、英語は日本語に、それぞれ翻訳される）、その翻訳結果としてのテキスト、その他付随する情報を、インターフェース部4や、出力部5、エージェント部6に出力する。

【0024】入力部3には、例えば、ユーザが発話した音声が入力されるようになっており、入力部3は、その音声を、電気信号としての音声データに変換し、音声認識部1やエージェント部6に出力する。

【0025】インターフェース部4は、音声認識部1と機械翻訳部2との間のインターフェースとして機能し、それらの間でのデータのやりとりを制御する。また、イ

## 5

ンターフェース部 4 は、必要な情報を、音声認識部 1 や機械翻訳部 2 から取得し、音声認識部 1 や、機械翻訳部 2、エージェント部 5 に出力する。

【0026】出力部 5 は、機械翻訳部 2 が出力する翻訳結果を、エージェント部 6 に出力したり、また、その翻訳結果を、合成音で出力し、あるいはテキスト等で表示する。

【0027】エージェント部 6 は、音声認識部 1 や、機械翻訳部 2、入力部 3、インターフェース部 4、出力部 5 から得られる情報に基づいて、例えば、画像として表示されるキャラクタや、実際の物体としてのロボット等のオブジェクトの状態を制御する。

【0028】以上のように構成される音声翻訳システムにおいては、例えば、日本語の音声が入力されると、その音声は、入力部 3 を介して音声認識部 1 に供給される。音声認識部 1 では、その音声は音声認識され、インターフェース部 4 を介して、機械翻訳部 2 に供給される。機械翻訳部 2 では、音声認識部 1 による日本語の音声認識結果が、英語に機械翻訳され、出力部 5 に供給される。出力部 5 では、機械翻訳部 2 による翻訳結果に対応する合成音が生成されて出力され、あるいは、その翻訳結果が表示される。

【0029】また、英語の音声が入力されると、その音声も、上述した場合と同様に、入力部 3 を介して音声認識部 1 に供給される。音声認識部 1 では、入力部 3 からの音声は音声認識され、インターフェース部 4 を介して機械翻訳部 2 に供給される。機械翻訳部 2 では、音声認識部 1 による英語の音声認識結果が、日本語に機械翻訳され、出力部 5 に供給される。出力部 5 は、機械翻訳部 2 による翻訳結果に対応する合成音が生成されて出力され、あるいは、その翻訳結果が表示される。

【0030】従って、図 1 の音声翻訳システムによれば、英語のユーザは、日本語のユーザによる日本語の発話を理解し、また、日本語のユーザは、英語のユーザによる英語の発話を理解することができ、日本語のユーザと、英語のユーザとの間で対話を行うことができる。

【0031】なお、以上のようにして、日本語のユーザと、英語のユーザとの間で対話が行われている場合において、エージェント部 6 は、音声認識部 1 や、機械翻訳部 2、入力部 3、インターフェース部 4、出力部 5 から得られる情報に基づいて、オブジェクトの状態を制御し、これにより、ユーザどうしの円滑なコミュニケーションを補助するようになっている。

【0032】次に、図 2 は、図 1 の音声認識部 1 および入力部 3 の構成例を示している。

【0033】同図に示すように、音声認識部 1 は、特徴抽出部 13、特徴量バッファ 14、マッチング部 15、音響モデルデータベース 16、辞書データベース 17、および文法データベース 18 で構成され、入力部 3 は、マイク 11 および AD 変換部 12 で構成されている。

## 6

【0034】ユーザの発話は、マイク 11 に入力され、マイク 11 では、その発話が、電気信号としての音声信号に変換される。この音声信号は、AD (Analog Digital) 変換部 12 に供給される。AD 変換部 12 では、マイク 11 からのアナログ信号である音声信号がサンプリング、量子化され、ディジタル信号である音声データに変換される。この音声データは、特徴抽出部 13 に供給される。

【0035】特徴抽出部 13 は、AD 変換部 12 からの音声データについて、適当なフレームごとに、例えば、スペクトルや、パワー、線形予測係数、ケプストラム係数、線スペクトル対等の特徴パラメータを抽出し、特徴量バッファ 14 およびマッチング部 15 に供給する。特徴量バッファ 14 では、特徴抽出部 13 からの特徴パラメータが一時記憶される。

【0036】マッチング部 15 は、特徴量抽出部 13 からの特徴パラメータ、または特徴量バッファ 14 に記憶された特徴パラメータに基づき、音響モデルデータベース 16、辞書データベース 17、および文法データベース 18 を必要に応じて参照しながら、マイク 11 に入力された音声（入力音声）を認識する。

【0037】即ち、音響モデルデータベース 16 は、音声認識する音声の言語における個々の音素や音節などの音響的な特徴を表す音響モデルを記憶している。ここで、音響モデルとしては、例えば、HMM (Hidden Markov Model) などを用いることができる。辞書データベース 17 は、認識対象の各単語（語句）について、その発音に関する情報が記述された単語辞書や、音素や音節の連鎖関係を記述した言語モデルを記憶している。文法データベース 18 は、辞書データベース 17 の単語辞書に登録されている各単語が、どのように連鎖する（つながる）かを記述した文法規則を記憶している。ここで、文法規則としては、例えば、文脈自由文法 (CFG) や、統計的な単語連鎖確率 (N-gram) などに基づく規則を用いることができる。

【0038】マッチング部 15 は、辞書データベース 17 の単語辞書、言語モデルを参照することにより、音響モデルデータベース 16 に記憶されている音響モデルを接続することで、単語の音響モデル（単語モデル）を構成する。さらに、マッチング部 15 は、幾つかの単語モデルを、文法データベース 18 に記憶された文法規則を参照することにより接続し、そのようにして接続された単語モデルを用いて、特徴パラメータに基づき、例えば、HMM 法等によって、マイク 11 に入力された音声

を認識する。

【0039】そして、マッチング部 15 による音声認識結果は、例えば、テキスト等で出力される。

【0040】なお、マッチング部 15 は、入力された音声を対象として、再度、処理を行うことが必要な場合は、特徴量バッファ 14 に記憶された特徴パラメータを

用いて処理を行うようになっており、これにより、ユーザに再度の発話を要求せずに済むようになっている。

【0041】次に、図3は、図1の機械翻訳部2の構成例を示している。

【0042】テキスト解析部21には、音声認識部1が出力する音声認識結果としてのテキスト等が、機械翻訳の対象として入力されるようになっており、テキスト解析部21は、辞書データベース24や解析用文法データベース25を参照しながら、そのテキストを解析する。

【0043】即ち、辞書データベース24には、各単語の表記や、解析用文法を適用するために必要な品詞情報などが記述された単語辞書が記憶されている。また、解析用文法データベース25には、単語辞書に記述された各単語の情報に基づいて、単語連鎖に関する制約等が記述された解析用文法規則が記憶されている。そして、テキスト解析部21は、その単語辞書や解析用文法規則に基づいて、そこに入力されるテキスト（入力テキスト）の形態素解析や、構文解析等を行い、その入力テキストを構成する単語や構文の情報等の言語情報を抽出する。ここで、テキスト解析部21における解析方法としては、例えば、正規文法や、文脈自由文法、統計的な単語連鎖確率を用いたものなどがある。

【0044】テキスト解析部21で得られた入力テキストの解析結果としての言語情報は、言語変換部22に供給される。言語変換部22は、言語変換データベース26を参照し、入力テキストの言語の言語情報を、翻訳結果の言語の言語情報に変換する。

【0045】即ち、言語変換データベース26には、入力言語（言語変換部22への入力の言語）の言語情報から、出力言語（言語変換部22からの出力の言語）の言語情報への変換パターン（テンプレート）や、入力言語と出力言語との対訳用例およびその対訳用例と入力言語との間の類似度の計算に用いられるシソーラス等の、言語情報を変換するための言語変換データが記憶されている。そして、言語変換部22では、このような言語変換データに基づいて、入力テキストの言語の言語情報が、出力言語の言語情報に変換される。

【0046】言語変換部22で得られた出力言語の言語情報は、テキスト生成部23に供給され、テキスト生成部23は、辞書データベース27および生成用文法データベース28を参照することにより、出力言語の言語情報から、入力テキストを出力言語に翻訳したテキストを生成する。

【0047】即ち、辞書データベース27には、出力言語の文を生成するのに必要な単語の品詞や活用形等の情報が記述された単語辞書が記憶されており、また、生成用文法データベース28には、出力言語の文を生成するのに必要な単語の活用規則や語順の制約等の生成用文法規則が記憶されている。そして、テキスト生成部23は、これらの単語辞書および生成用文法規則に基づい

て、言語変換部22からの言語情報を、テキストに変換して出力する。

【0048】次に、図4は、図1の出力部5の音声合成処理を行う部分の構成例を示している。

【0049】テキスト解析部31には、機械翻訳部2が出力する翻訳結果としてのテキストテキストが、音声合成処理の対象として入力されるようになっており、テキスト解析部31は、辞書データベース34や解析用文法データベース35を参照しながら、そのテキストを解析する。

【0050】即ち、辞書データベース34には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、解析用文法データベース35には、辞書データベース34の単語辞書に記述された単語について、単語連鎖に関する制約等の解析用文法規則が記憶されている。そして、テキスト解析部31は、この単語辞書および解析用文法規則に基づいて、そこに入力されるテキストの形態素解析や構文解析等の解析を行い、後段の規則合成部32で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを制御するための情報その他の韻律情報や、各単語の発音等の音韻情報などがある。

【0051】テキスト解析部31で得られた情報は、規則合成部32に供給され、規則合成部32では、音素片データベース36を用いて、テキスト解析部31に入力されたテキストに対応する合成音の音声データ（デジタルデータ）が生成される。

【0052】即ち、音素片データベース36には、例えば、CV(Consonant, Vowel)や、VCV、CVC等の形で音素片データが記憶されており、規則合成部32は、テキスト解析部31からの情報に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を適切に付加することで、テキスト解析部31に入力されたテキストに対応する合成音の音声データを生成する。

【0053】この音声データは、DA変換部33に供給され、そこで、アナログ信号としての音声信号に変換される。この音声信号は、スピーカ37に供給され、これにより、テキスト解析部31に入力されたテキストに対応する合成音が出力される。

【0054】次に、図5および図6を参照して、インターフェース部4の処理について説明する。

【0055】インターフェース部4は、図5に示すように、音声認識部1が出力する音声認識結果を受信し、そのフォーマットを、機械翻訳部2で取り扱い可能なフォーマットに変換して、機械翻訳部2に供給する。

【0056】また、インターフェース部4は、図6に示すように、バッファ4Aおよび情報交換部4Bを有して

いる。そして、インターフェース部4は、音声認識部1による音声認識処理の途中結果等の内部動作状態を取得し、バッファ4Aに記憶させるとともに、機械翻訳部2による機械翻訳処理の途中結果等の内部動作状態を取得し、バッファ4Aに記憶させる。さらに、インターフェース部4では、情報交換部4Bにおいて、バッファ4Aに記憶された音声認識部1の内部動作状態が、必要に応じて、機械翻訳部2にフィードフォワードされるとともに、バッファ4Aに記憶された機械翻訳部2の内部動作状態が、必要に応じて、音声認識部1にフィードバックされる。この場合、機械翻訳部2は、インターフェース部4からフィードフォワードされる音声認識部1の情報をを用いて処理を行うことにより、翻訳精度を向上させる。同様に、音声認識部1も、インターフェース部4からフィードバックされる機械翻訳部2の情報をを用いて処理を行うことにより、音声の認識精度を向上させる。

【0057】次に、エージェント部6は、音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5から得られる情報に基づいて、オブジェクトの状態を制御し、これにより、例えば、発話を行っているユーザの感情や、音声翻訳システムの内部状態等の各種の情報を、ユーザに提示することで、ユーザどうしが円滑なコミュニケーションを図るための、いわば手助けを行うようになっている。

【0058】即ち、エージェント部6は、図7のフローチャートに示すように、まず最初に、ステップS1において、音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5から何らかの情報が入力されたかどうかを判定する。ステップS1において、音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5から情報が入力されていないと判定された場合、ステップS1に戻る。

【0059】また、ステップS1において、音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5から情報が入力された場合、ステップS2に進み、その情報に基づいて、オブジェクトの状態を変化させる。そして、ステップS1に戻り、以下、同様の処理が繰り返される。

【0060】音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5から、エージェント部6に対しては、例えば、図8に示すような情報を入力することができる。

【0061】即ち、音声認識部1からは、例えば、入力された音声の言語的な特徴（言語特徴）や、その音声を発話したユーザの特徴（話者特徴）等を表す情報が、エージェント部6に入力される。ここで、入力された音声の言語特徴を表す情報としては、例えば、その音声の言語の特徴（例えば、日本語による音声が入力された場合は、日本語の特徴）等がある。ユーザの話者特徴を表す情報としては、例えば、入力された音声の信号波形や、

スペクトラム、フォルマント周波数、ピッチ周波数等がある。

【0062】さらに、音声認識部1からエージェント部6に対しては、例えば、その内部動作状態を表す情報を入力することも可能である。ここで、音声認識部1の内部動作状態を表す情報としては、例えば、最終的な音声認識結果や、その音声認識結果に関連する画像データ、音声認識結果の候補、音声認識処理の進行の程度等がある。

10 【0063】また、機械翻訳部2からは、例えば、翻訳結果の言語的な特徴（言語特徴）や、その意味内容等を表す情報が、エージェント部6に入力される。ここで、翻訳結果の言語特徴を表す情報としては、例えば、その翻訳結果の言語の特徴（例えば、英語による翻訳結果の場合は、英語の特徴）等がある。翻訳結果の意味内容を表す情報としては、例えば、その翻訳結果のテキスト等がある。

20 【0064】さらに、機械翻訳部2からエージェント部6に対しては、例えば、その内部状態を表す情報を入力することも可能である。ここで、機械翻訳部2の内部状態を表す情報としては、例えば、機械翻訳の進行の程度、最終的な翻訳結果、翻訳結果の候補等がある。

【0065】また、インターフェース部4からは、例えば、その内部状態を表す情報が、エージェント部6に入力される。ここで、インターフェース部4の内部状態を表す情報としては、例えば、バッファ4A（図6）に記憶された情報や、インターフェース部4の処理状態等がある。

30 【0066】エージェント部6は、上述したような情報に基づいて、オブジェクトの、例えばパフォーマンス等を変化させる。

【0067】即ち、エージェント部6は、例えば、図9に示すように、音声認識部1による音声認識結果や、機械翻訳部2による翻訳結果に、悲しい意味の言葉（語句）が含まれる場合には、悲しいようなパフォーマンスのオブジェクト（悲しそうな音声を出力したり、悲しそうな動作を行うオブジェクト）#1を提示し、嬉しい言葉が含まれる場合には、嬉しそうなパフォーマンスのオブジェクト#2を提示する。

40 【0068】また、エージェント部6は、例えば、音声認識部1が認識する音声の言語的な特徴や、機械翻訳部2が出力する翻訳結果の言語的な特徴に基づいて、オブジェクトのパフォーマンスを変化させる。即ち、例えば、日本語には、比較的曖昧な特徴があり、英語には、比較的フランクな特徴がある。そこで、エージェント部6は、例えば、図10に示すように、日本語による翻訳文を出力する場合には、曖昧さを有するパフォーマンスのオブジェクトを提示し、英語による翻訳文を出力する場合には、フランクなパフォーマンスのオブジェクトを提示する。さらに、エージェント部6は、例えば、中国



語による翻訳文を出力する場合には、中国語の言語的な特徴を有するパフォーマンスのオブジェクトを提示する。

【0069】なお、エージェント部6において、複数のオブジェクトを扱うことが可能な場合には、その複数のオブジェクトそれぞれを、異なるパフォーマンスを有するものとして提示することができる。また、エージェント部6において、1つのオブジェクトしか扱うことが可能でない場合には、異なるパフォーマンスの提示は、その1つのオブジェクトのパフォーマンスを、適宜変更することで行うことができる。

【0070】また、オブジェクトには、例えば、音声認識部1による音声認識結果や、機械翻訳部2による翻訳結果を、手話で表現させることができる。この場合、音声翻訳システムを、聴覚に障害があるユーザとのコミュニケーションツールとしても用いることが可能となる。

【0071】オブジェクトとしては、上述したように、画像として表示されるキャラクタ（例えば、ミッキーマウスや、キティ（いずれも商標）等のようなキャラクタ）や、実際の物体としてのロボット（例えば、AIBOやファジー（いずれも商標）等のようなロボット）等を用いることが可能である。また、オブジェクトとして、画像として表示されるキャラクタを用いる場合には、そのキャラクタが発話をしているかのような音声（合成音）を出力するようにすることが可能である。

【0072】なお、オブジェクトとしては、その他、音声（合成音）を用いることも可能である。オブジェクトとして、音声を用いる場合には、オブジェクトは、ユーザにとって声優のような存在となり、エージェント部6では、そのオブジェクトとしての音声のピッチ周波数等が、音声認識部1や、機械翻訳部2、入力部3、インターフェース部4、出力部5から得られる情報に基づいて制御される。

【0073】オブジェクトとして、画像として表示されるキャラクタを用いる場合には、そのキャラクタを表示するためのハードウェア（例えば、液晶モニタなど）およびソフトウェアが必要となる。また、オブジェクトとして、音声を用いる場合には、その音声を出力するためのハードウェア（例えば、スピーカなど）およびソフトウェアが必要となる。さらに、オブジェクトとして、ロボットを用いる場合には、そのロボットを実現するためのハードウェアおよびソフトウェアが必要となる。

【0074】オブジェクトとして用いるハードウェア（ソフトウェアも同様）は、図1の音声翻訳システムと一体的に構成することもできるし、別個独立に構成することもできる。但し、別個独立に構成する場合には、オブジェクトとしてのハードウェアと、図1の音声翻訳システムとの間で通信を行うことができるようにする必要がある。

【0075】オブジェクトとして、ロボットを用いて、

そのロボットと音声翻訳システムとを一体的に構成したシステムの構成例を、図11および図12に示す。

【0076】図11は、オブジェクトとして、1体のロボットを採用した場合を示しており、そのロボットには、図1に示した音声翻訳システムが内蔵されている。なお、図11においては、入力部3を構成するAD変換部12（図2）を、信号処理部と記述しており、また、出力部5を構成するテキスト解析部31、規則合成部32、DA変換部33、辞書データベース34、解析用文法データベース35、および音楽片データベース36（図4）を、まとめて音声合成部と記述してある。

【0077】図12は、オブジェクトとして、2対のロボット51Aおよび51Bを採用した場合を示している。

【0078】ロボット51Aは、図1の音声翻訳システムと同様に構成される音声翻訳部52Aを内蔵しており、ロボット51Bも、図1の音声翻訳システムと同様に構成される音声翻訳部52Bを内蔵している。

【0079】但し、図12の実施の形態では、ロボット51Aが内蔵する音声翻訳部52Aを構成するインターフェース部4（図1）と、ロボット51Bが内蔵する音声翻訳部52Bを構成するインターフェース部4との間で通信を行うことができるようになっている。これにより、例えば、ロボット51Aのマイク11に入力された音声を、その音声翻訳部52Aで音声認識し、その音声認識結果を、ロボット51Bの音声翻訳部52Bに送信することができるようになっている。そして、ロボット51Bの音声翻訳部52Bで、ロボット51Aからの音声認識結果を翻訳し、その翻訳結果に対応する合成音を、ロボット51Bのスピーカ37から出力することができるようになっている。

【0080】次に、エージェント部6は、音声認識部1による音声認識結果や、機械翻訳部2による翻訳結果の補正を行うことができるようになっている。

【0081】即ち、例えば、音声認識部1は、音声認識を行うことにより得られる音声認識結果の候補の中から、その尤度の最も高いものを最終的な音声認識結果として出力するようになっており、同様に、機械翻訳部2も、翻訳結果の候補の中から、その尤度の最も高いものを最終的な翻訳結果として出力するようになっているが、例えば、最高の尤度と、次に高い尤度との間に、ほとんど差がない場合には、最高の尤度を有する候補を、最終的な音声認識結果や翻訳結果とすることが必ずしも適切であるとはいえない。

【0082】そこで、エージェント部6は、音声認識部1において得られた音声認識結果の候補どうしや、機械翻訳部2の翻訳結果の候補どうしの尤度に大きな差がない場合には、上位の尤度を有する幾つかの候補を、オブジェクトを通じて提示し（例えば、表示し）、ユーザに、その中から、確からしいものを選択してもらう。そ

して、音声認識部 1、または機械翻訳部 2 は、選択された候補によって、最終的な音声認識結果、または翻訳結果を補正する。なお、音声認識の分野では、このような補正方法は、コンフィデンスメジャー (Confidence Measure) と呼ばれる。

【0083】なお、エージェント部 6 には、上位の尤度を有する幾つかの候補の共通点を、オブジェクトを通じて提示させるようにすることも可能である。

【0084】次に、音声認識部 1 (図 2) における音響モデルデータベース 16 や、辞書データベース 17、文法データベース 18 に記憶された情報 (以下、適宜、音声認識部 1 の知識ベースという) は、いわゆるモデルのオンライン適応等を行うことで、より認識性能を向上させるように変更 (更新) することができる。同様に、機械翻訳部 2 (図 3) における辞書データベース 24 や、解析用文法データベース 25、言語変換データベース 26、辞書データベース 27、生成用文法データベース 28 に記憶された情報 (以下、適宜、機械翻訳部 2 の知識ベースという) も、同様に、より翻訳性能を向上させるように変更することができる。

【0085】そして、この場合、エージェント部 6 には、上述のような音声認識部 1 や機械翻訳部 2 の知識ベースの変更に基づいて、オブジェクトの状態を変更させるようにすることができる。

【0086】さらに、オブジェクトを生命体として捉え、エージェント部 6 には、音声認識部 1 や機械翻訳部 2 の知識ベースの変更に基づいて、オブジェクトを、生物と同様に成長させることができる。

【0087】即ち、例えば、いま、オブジェクトとして、画像として表示されるキャラクタを採用することとして、図 13 に示すように、エージェント部 6 には、音声認識部 1 や機械翻訳部 2 の知識ベースが入力される。エージェント部 6 では、この知識ベースに基づいて学習が行われ、例えば、言語能力が向上し、表現の豊かなキャラクタが提示される。

【0088】また、エージェント部 6 では、知識ベースの程度に基づいて、キャラクタが、生物と同様に、新生児から、幼児、少年、青年、中年、老年、さらに高年の状態を表す画像に変更される。さらに、この場合、エージェント部 6 では、キャラクタの成長の度合いに応じて、その行動等も変更される。即ち、例えば、年齢が高くなるほど、疲れの表情をよく見せるようになるキャラクタ等に変更される。

【0089】なお、上述の場合には、知識ベースの程度に基づいて、キャラクタの状態を変更するようにしたが、キャラクタの状態は、その他、例えば、音声翻訳システムの使用時間に基づいて変更するようにすることも可能である。この場合、キャラクタの成長の度合いに応じて、その行動や知識等をも変更するようにすることが可能である。

【0090】また、例えば、音声翻訳システムが、携帯型で、バッテリーを電源として動作するものである場合には、キャラクタの状態は、そのバッテリーの残存容量に基づいて、上述したように変更するようにすることが可能である。

【0091】次に、エージェント部 6 には、音声認識部 1 や、機械翻訳部 2、入力部 3、インターフェース部 4、出力部 5 からのユーザの感情を表す情報に基づいて、その感情を表現するキャラクタ (あるいは、その他のオブジェクト) を提示させるようにすることが可能である。

【0092】ここで、ユーザの感情を表す情報としては、例えば、図 14 に示すように、発話者の表情や、発話文章の内容に含まれる感情成分、翻訳された文章の感情成分、相手の表情等がある。

【0093】発話者の表情や相手の表情は、例えば、入力部 3 に、CCD (Charge Coupled Device) カメラを内蔵させ、ユーザを撮影することによって取得することが可能である。

【0094】また、発話文章の内容に含まれる感情成分としては、ユーザが発話した音声の音響的な特徴としての、例えば、音声波形の変化やフォルマントの変化等があり、これらは、例えば、音声認識部 1 の特徴抽出部 13 において音響分析を行うことにより取得することが可能である。

【0095】さらに、発話文章の内容に含まれる感情成分としては、ユーザが発話した音声の音声認識結果の言語的な特徴としての、例えば、その音声認識結果に含まれる言葉 (単語) の意味や、前後の音声認識結果の内容の論理関係、音声認識結果の文法関係等があり、これらは、例えば、機械翻訳部 2 のテキスト解析部 21 においてテキスト解析を行うことにより取得することが可能である。

【0096】また、翻訳された文章の感情成分としては、ユーザが発話した音声を翻訳した翻訳結果の言語的な特徴としての、例えば、その翻訳結果に含まれる言葉の意味や、前後の翻訳結果の内容の論理関係、翻訳結果の文法関係等があり、これらは、例えば、機械翻訳部 2 が出力する翻訳結果を解析することにより取得することが可能である。

【0097】ここで、本実施の形態では、音声翻訳システムにおいて、日英と英日の両方の翻訳が可能であり、上述したように、日本語のユーザと、英語のユーザとの間で対話を行うことができるが、エージェント部 6 には、この対話の流れから、それぞれのユーザの感情を認識させ、その感情を表現するキャラクタ (あるいは、その他のオブジェクト) を提示させることも可能である。

【0098】以上のように、キャラクタ等に、ユーザの感情を表現させることで、ユーザの意図を、相手のユーザに正確に伝えることが可能となり、その結果、円滑な

コミュニケーションを図ることが可能となる。

【0099】なお、その他、図14に、点線で囲って示すように、エージェント部6には、音声翻訳システムに特有な内部状態に基づいて、感情を変化させるキャラクタ（あるいは、その他のオブジェクト）を提示させるようにすることが可能である。即ち、例えば、音声翻訳システムが、携帯型で、バッテリーを電源として動作するものである場合には、そのバッテリーの残存容量に基づいて、キャラクタの感情表現を変化させることが可能である。さらに、例えば、音声翻訳システムにおいて有効になっている機能（あるいは無効になっている機能や故障箇所等）に基づいて、キャラクタの感情表現を変化させることも可能である。

【0100】次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0101】そこで、図15は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0102】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0103】あるいはまた、プログラムは、フロッピーディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体111に、一時的あるいは永続的に格納（記録）しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。

【0104】なお、プログラムは、上述したようなリムーバブル記録媒体111からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部108で受信し、内蔵するハードディスク105にインストールすることができる。

【0105】コンピュータは、CPU(Central Processing Unit)102を内蔵している。CPU102には、バス101を介して、入出力インタフェース110が接続されており、CPU102は、入出力インタフェース110を介して、ユーザによって、キーボードやマウス等で構成される入力部107が操作されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納

されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM(Random Access Memory)104にロードして実行する。これにより、CPU102は、上述した各種の処理を実行する。そして、CPU102は、その処理結果を、必要に応じて、例えば、入出力インタフェース110を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0106】ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0107】また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さらに、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0108】

【発明の効果】本発明の情報処理装置および情報処理方法、並びに記録媒体によれば、音声翻訳装置から得られる情報に基づいて、制御対象であるオブジェクトの状態が制御される。従って、例えば、オブジェクトに、ユーザの感情を表現させることで、ユーザの意図を、相手のユーザに正確に伝えることが可能となり、その結果、円滑なコミュニケーションを図ることが可能となる。

【図面の簡単な説明】

【図1】本発明を適用した音声翻訳システムの一実施の形態の構成例を示すブロック図である。

【図2】音声認識部1および入力部3の構成例を示すブロック図である。

【図3】機械翻訳部2の構成例を示すブロック図である。

【図4】出力部5の構成例を示すブロック図である。

【図5】インターフェース部4を説明するための図である。

【図6】インターフェース部4を説明するための図である。

【図7】エージェント部6の動作を説明するためのフローチャートである。

【図8】エージェント部6への入力を説明するための図である。

【図9】エージェント部6を説明するための図である。

【図10】オブジェクトの例を示す図である。

17

【図11】オブジェクトとしてのロボットを示す図である。

【図12】オブジェクトとしてのロボットを示す図である。

【図13】エージェント部6を説明するための図である。

【図14】エージェント部6への入力を説明するための図である。

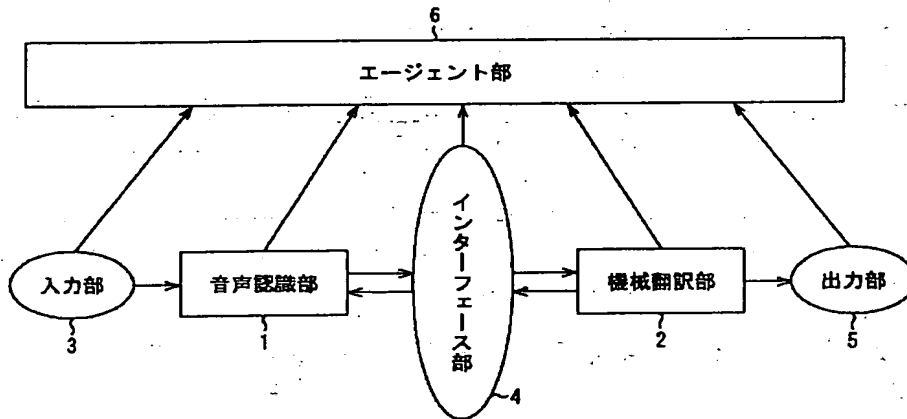
【図15】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

1 音声認識部, 2 機械翻訳部, 3 入力部,  
4 インターフェース部, 4A バッファ, 4B  
情報交換部, 5 出力部, 6 エージェント部,  
11 マイク (マイクロフォン), 12 AD変換  
部, 13 特徴抽出部, 14 特徴量バッファ,

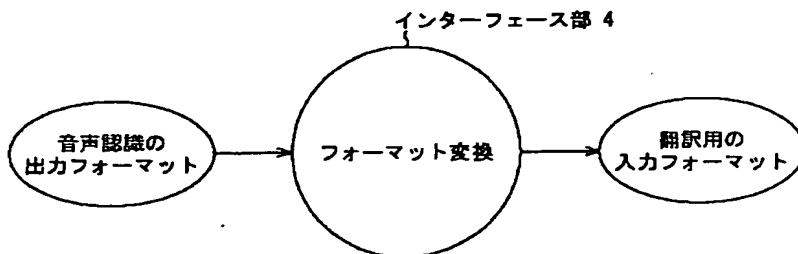
15 マッチング部, 16 音響モデルデータベ  
ース, 17 辞書データベース, 18 文法データ  
ベース, 21 テキスト解析部, 22 言語変換部,  
23 テキスト生成部, 24 辞書データベ  
ース, 25 解析用文法データベース, 26 言語変換デー  
タベース, 27 辞書データベース, 28 生成用  
文法データベース, 31 テキスト解析部, 32 規  
則合成部, 33 DA変換部, 34 辞書データベ  
ース, 35 解析用文法データベース, 36 音素  
片データベース, 37 スピーカ, 51A, 51B  
ロボット, 52A, 52B 音声翻訳部, 101  
バス, 102 CPU, 103 ROM, 104 RA  
M, 105 ハードディスク, 106 出力部, 1  
07 入力部, 108 通信部, 109 ドライブ,  
110 入出力インタフェース, 111 リムーバ  
ブル記録媒体

【図1】



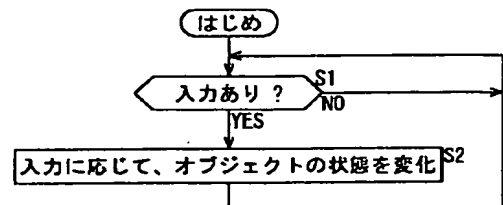
音声翻訳システム

【図5】

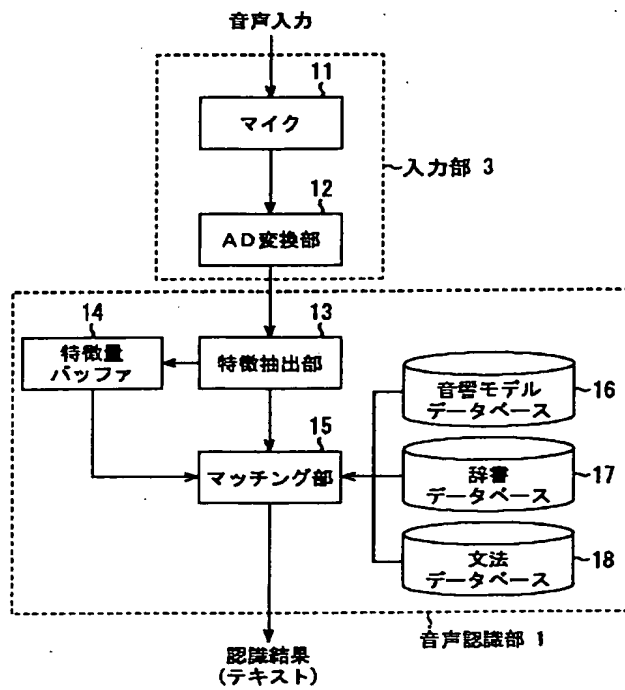


インターフェース部によるフォーマット変換

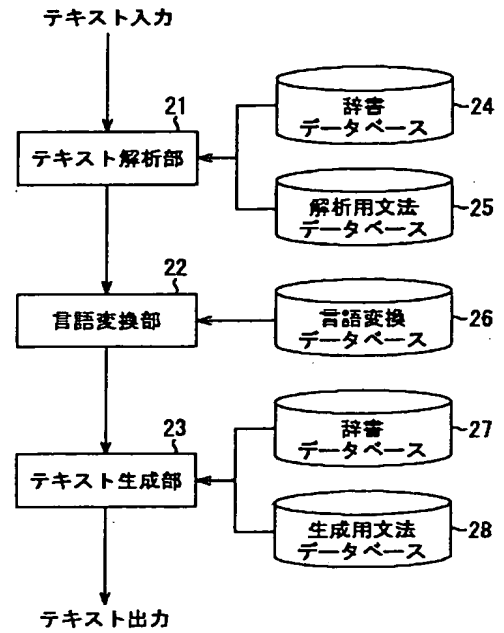
【図7】



【図 2】

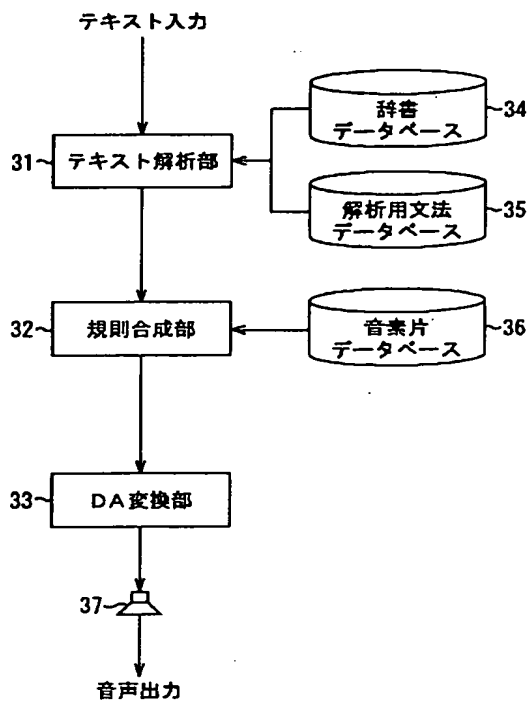


【図 3】



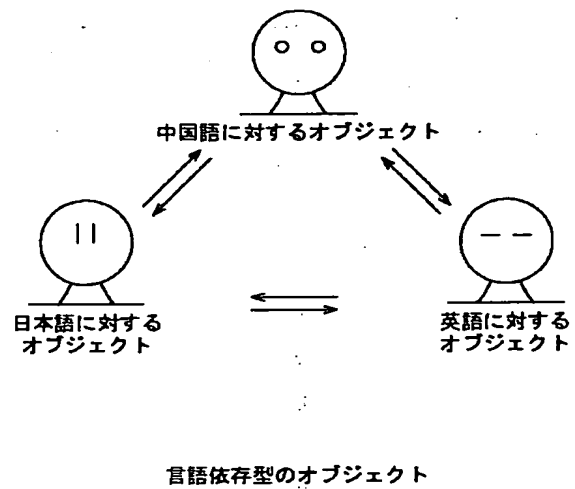
機械翻訳部 2

【図 4】

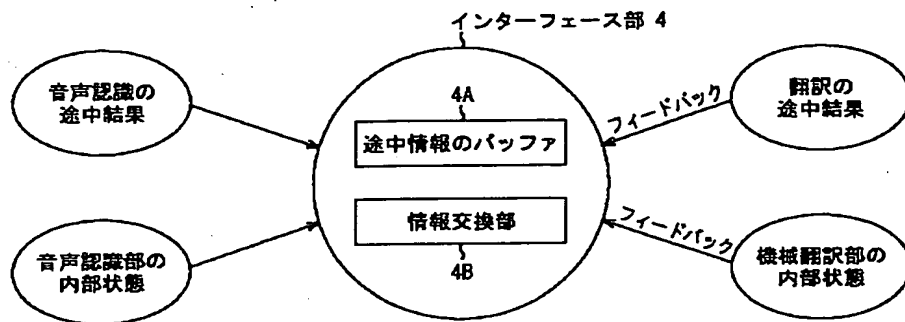


出力部 5

【図 10】

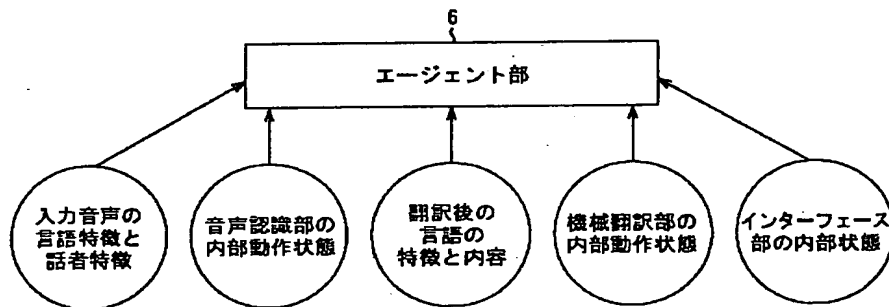


【図 6】



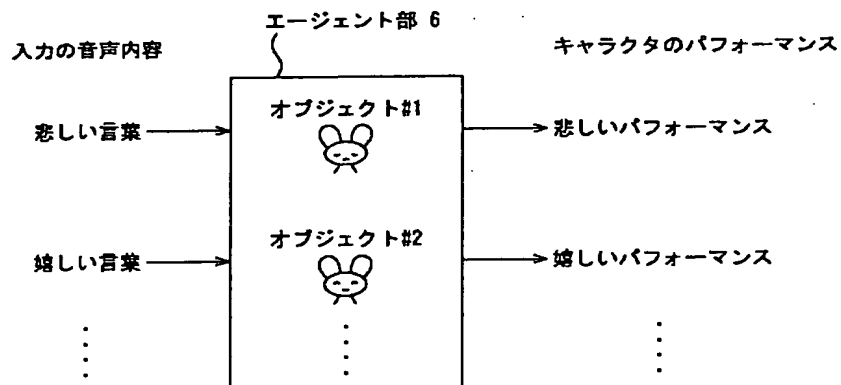
インターフェース部による音声認識部と機械翻訳部との間の瞬時情報交換

【図 8】



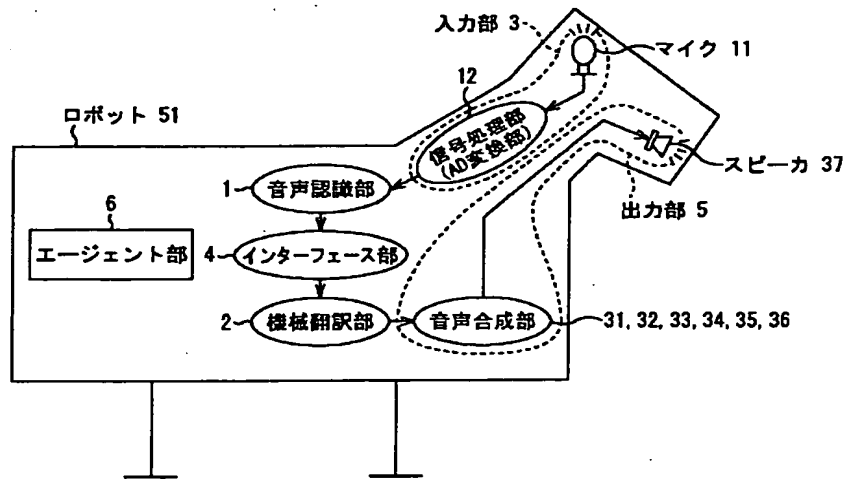
エージェント部による外部への呈示の一例

【図 9】



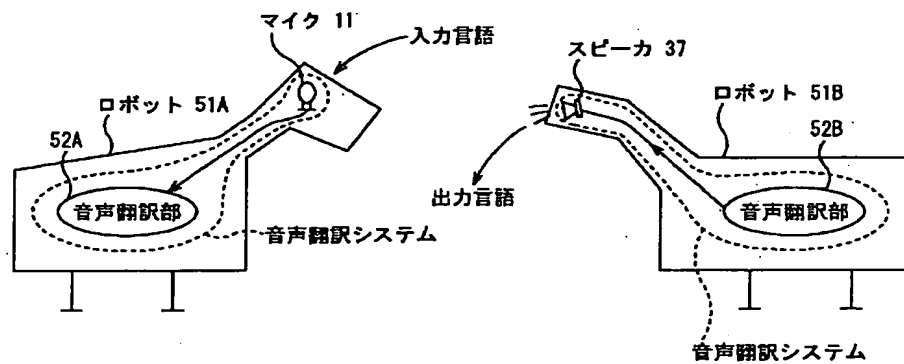
オブジェクトパフォーマンスの変化

【図11】



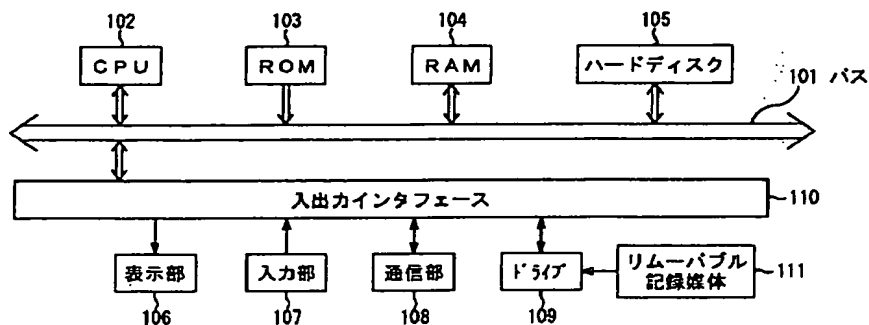
オブジェクトをロボットにした場合(単体)

【図12】



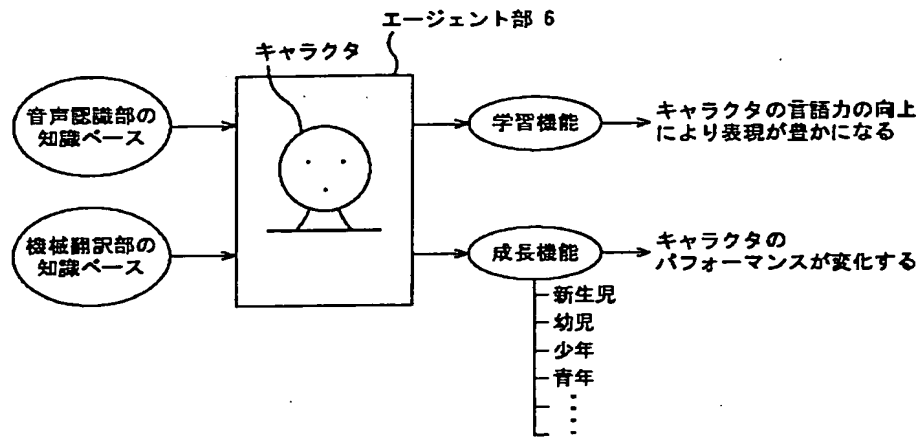
オブジェクトをロボットにした場合(分離、対話型)

【図15】



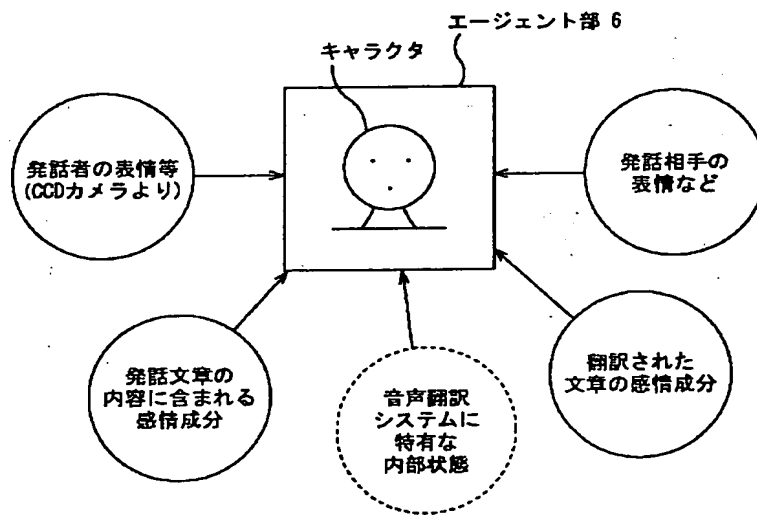
コンピュータ

【図 13】



知識ベースに依存するキャラクターの学習と成長

【図 14】



キャラクターの感情表現